DOCUMENT RESUME

ED 347 190 TM 018 667

AUTHOR Dean, Paul; And Others
TITLE Studying School Improvement in Chicago: A Construction of Post-Reform School Index Measures.
INSTITUTION Chicago Panel on Public School Policy and Finance, IL.; Chicago Public Schools, Ill.; Chicago Univ., Ill. Dept. of Education.
SPONS AGENCY Spencer Foundation, Chicago, Ill.
PUB DATE Apr 92
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Comparative Analysis; Educational Assessment; Educational Change; *Educational Improvement; Elementary Education; *Elementary School Students; Mathematical Models; Measurement Techniques; *School Districts; Scores; *Standardized Tests; Test Construction; Test Results; Urban Schools
IDENTIFIERS *Chicago Public Schools IL; Iowa Tests of Basic Skills; *Reform Efforts

ABSTRACT

This paper is the second report on the purpose, methodology, and results of a large-scale statistical analysis of standardized test scores from Chicago public elementary schools for 1988 and 1989. Major purposes were to develop methods for measuring student improvement (change) at the individual and school levels, to find ways of identifying improving schools, and to develop a better knowledge of the correlates of student and school improvement, especially with respect to school reform under the Chicago School Reform Act. Large-scale databases of student test scores are being used, including scores on the Iowa Tests of Basic Skills. Analytic issues included the metric to be used in modeling growth, the accuracy of any assessment of change, the problem of student mobility, and a way of reporting growth results at the school level. Predicted gains to be used in deriving school improvement indices were calculated for norming samples of between 5,800 and 6,300 students at each grade level. Actual results for four schools were examined in the construction of growth indices. The next step will be establishing a more reliable 2-year baseline and index scores for particular programs. Correlates of school improvement will be further defined, and additional student level data will be obtained. Appendix A contains 14 bar graphs. Appendix B and Appendix C contain eight tables of analysis results. There is a 10-item list of references. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

PAUL DEAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED347190

Studying School Improvement in Chicago:

A Construction of Post-Reform School Index Measures

Paul Dean
University of Chicago

Anthony S. Bryk
University of Chicago

John Q. Easton
Chicago Panel on Public School Policy and Finance

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1992.

TM018667

BEST COPY AVAILABLE

Studying School Improvement in Chicago:

A Construction of Post-Reform School Index Measures*

Introduction

This paper is the second report covering the purpose, methodology, and results of a large scale statistical analysis of standardized test scores from Chicago public elementary schools for the years 1988 and 1989 (see Easton, Dean, and Bryk, 1991). We will show our progress to date on this project and discuss many of the methodological and substantive issues that have arisen as a result of our work.

Purpose

The major purpose of this study is to develop methods for adequately measuring student improvement (change) at the individual and school level, to find ways of identifying improving schools, and to develop a better knowledge of the correlates of student and school improvement, especially with respect to school reform. The goals of the Chicago School Reform Act, which provided the impetus for this study, state that each school in the system will have at least 50 percent of its students scoring at or above the 50th percentile on national norms of standardized tests, and in reading and math, by 1993.

This is an inadequate long term goal for a major school improvement project. It does not provide an acceptable means of tracking schools to determine whether or not they are improving over time. Nor does it attempt to take into account many complex issues, at the student or school level, such as student background, mobility, and special program status. Therefore, we are attempting to develop a method to find out how student learning, as imperfectly measured by standardized tests, is changing over time.

In order to do this, we are working with large scale data bases of student test scores provided by the Chicago Public Schools.Because of their own interest in this work, and because of prior related analytic work, the school system has supplied us with computer tapes containing all Iowa Test of Basic Skills (ITBS) scores and item responses for all elementary school students for the spring 1988 and 1989 citywide test administrations, and have agreed to provide test results for subsequent years. The Chicago Public Schools has also made additional student level information available, binding us to the same level of confidentiality that they adhere to.

---

*This project is funded by a grant from the Spencer Foundation to the Chicago Panel on Public School Policy and Finance. It is a collaborative effort of the Chicago Panel, the Center for School Improvement in the Department of Education at the University of Chicago, and the Chicago Public Schools. The authors wish to express their thanks to Ong-Kim Lee, David Kerbow, Julie Smith, Ben Wright, Ken Frank and the many other individuals whose helpful comments have made this report possible, and to Arie van der Ploeg and Carole Perlman for their help in obtaining and working with the data.

Analytic Issues

Our first analytic consideration was the metric to be used in modeling growth over time. The Chicago Public Schools has traditionally reported student scores in grade equivalents (GEs). These scores have the benefit of being easily interpreted by teachers and parents. However, earlier research on Chicago Public Schools test results has shown that GEs are not adequate for modeling achievement growth (Schulz, Shen, and Wright, 1990), at least in part because they are intentionally designed to make gains equal for each year in school. Therefore, since we are specifically interested in studying gains, we have decided to use an alternative metric derived from the one parameter Rasch model (an Item Response Theory model).

Second, because the Chicago Public Schools are using different forms of the ITBS for five consecutive years (1989 through 1993) we were concerned about the accuracy of any assessment of change over time given possible equating problems within and between forms of the ITBS. Current scores may not be comparable to previous scores because some forms have never been directly equated. In addition, no equating studies have been conducted specifically in Chicago where instructional and sub-population differences might be significant. Therefore, we implemented an equating study, to be completed in sequence with the Chicago testing, to equate the five forms. The first part of the equating has been completed, (see Lee and Wright, 1992), and this report uses the Rasch measures from this equating to construct our 1988 to 1989 gain scores (i.e., our outcome measures).

Third, we needed to find a way to deal with student mobility, one of the most difficult issues that urban schools face. Many students, often those without strong preparation for school, move regularly from one school to another both during the school year, and between school years. This is a major practical problem for the schools themselves because it is clearly hard to educate students who move in and out of schools, and this is much more likely to happen in certain schools (see Kerbow, 1992). It is also a difficult one for researchers. Because student enrollment is essentially an administrative issue, school records on school movement are kept independently of test scores and special program enrollment records.

In Chicago, student test files do not account for student school changes. They simply indicate the school that the student was enrolled in at the time of the spring ITBS testing. From this information we are only able to tell if the student was or was not in the same school from one April testing to the next. Using this information we might miss, for example, if a student left that school and returned between the two testings, which is relatively common. We are also unable to model multiple moves that some students might have made, which could further explain changes in achievement growth. Other records of student mobility do exist, which we will discuss shortly. However, because of the enormity of the data we are working with (roughly 26,000 students per grade for grades one through eight), translating this data for our purposes has become a major obstacle.

Therefore, because we are not able to locate students in schools at precise times during the school year, we decided to drop from our analyses all students who were not in the same school for both the 1988 and 1989 April testings. We felt, as discussed above, that

it was unwise to hold a school "accountable" for any student that we did not believe was in that school for the probable majority of the 1989 school year. Exceptions to this dropping were, for example, schools that accepted students from only certain grade levels (i.e.,not the modal kindergarten through eighth grade schools) wherein all entering students were counted since all students in that school had moved, and moves to branch or feeder schools, for example, from a kindergarten through second grade school to its branch third through eighth grade school.

Finally, we needed to find a way of reporting our student growth results at the school level, one that we thought researchers and practitioners would find useful and easy to understand, especially given the complexity of our procedures. We have decided to develop four separate indices for each school to model student growth and school improvement over time. Each of the indices is set up in a fashion similar to the Consumer Price Index, where a score of 100 is designed as a reference point for each school in the base year of 1989 (the year just before the Chicago Reform Act went into effect). Comparisons are then made based on growth for that year, and therefore results should be broadly interpretable as a percent increase or decrease in growth as compared to the 1989 base year. In the future we will also use 1987 to 1988 gains in order to form a more stable two-year growth baseline. Therefore, researchers, school personnel, and policy makers are free to choose the index or indices that they believe are most valid for helping them with decisions that they face.

The School Improvement Indices

The school score for each of the four indices is calculated by taking the average actual mean gain for a single school, dividing this by the average predicted mean gain, and then multiplying by 100. The indices differ from each other in how the predicted gains are determined. Thus, the general form of the indices is:

$$Index = \frac{Mean\ Observed\ Gain}{Mean\ Predicted\ 89\ Gain} \times 100$$

We are calling our first index, shown below, the system index because it compares each school's growth against the whole system average growth in the base year 1989 (i.e., pre-reform), and does not control for any student level characteristics. The actual average system gains, for reading and math, are discussed in the Results section.

$$System\ Index = \frac{Mean\ Observed\ Gain}{Mean\ System\ 89\ Gain} \times 100$$

The second index, our school index, compares each school's growth in any year against that school's growth in the base year 1989, and like the system index, it does not control for any student level characteristics. No scores are being reported in this paper for the school index because, again, only 1988 to 1989 gains have been calculated to this point, and all schools have a value of exactly 100 (by definition) on this index in 1989.

$$School\ Index = \frac{Mean\ Observed\ Gain}{Mean\ School\ 89\ Gain} \times 100$$

Index 3 is our background index because the prediction equation that is used to form the denominator regresses the student's Rasch gain against their background characteristics of race, gender,and 'old for grade' (explained in the Methods section), and this individual predicted score is then averaged at the school level.

$$Background\ Index = \frac{Mean\ Observed\ Gain}{Mean\ 89\ Gain\ Predicted\ by\ Student\ Demographics} \times 100$$

We are calling index 4 the program index because it includes Chapter One program status, Bilingual program status, Special Education program status, and previous year retention as independent variables, in addition to those in index 3 above, in the regression equation used to find the predicted score. Reporting of program index scores is not included in this paper.

$$Program\ Index = \frac{Mean\ Observed\ Gain}{Mean\ 89\ Gain\ Predicted\ by\ Student\ Background\ plus\ Program\ Participation} \times 100$$

These indices are calculated for each grade level (two through eight) separately for reading and math. The whole school level index score is a weighted average of these grade level index scores**. Thus, the information that we hope to be providing to schools will be, for each year separately for reading and math, the number of students that passed through the cleanings at each grade level, the grade level index scores for each of the four indices above, the total number of students in the school that passed the cleanings,and the whole school index scores for each of the four indices. Some sample output of this type is analyzed in the Results section.

Methods

A simple random one-quarter sample of students was taken at each grade level. These norming samples of between 5800 and 6300 students (more at the lower grades, as expected), were taken in order to determine the predicted gains that were to be used as the denominator of the third and fourth indices. In calculating these predicted gains, we first went through many cleanings of the data. We attempted to follow as closely as possible the Chicago Public Schools' own criteria for deciding what students to include in their reporting. This includes, for example, cleanings for invalid fields of data, and the exclusion of many types of bilingual and special education students.In addition, we instituted many other cleaning steps. For example, we checked for matching 1988 and 1989 information (student

**Although more mathematically appropriate computations are available, we chose to use the weighted average for its ease of interpretation.

identification number, race, gender, etc.), and we only included those students who were in their proper cohort (that is, born after December 1st, as is the Chicago Public Schools' policy), or exactly one year behind their cohort (born after December 1st of the next year, but no later than November 30th of that following year). The latter students were labeled "old for grade," however, we cannot determine whether these students were retained or if they started school late. In addition, for example, we eliminated any students who had skipped a grade or moved back a grade from 1988 to 1989, and any students who took a test more than two levels higher or more than one level lower than in their previous testing.

Next, the mobility cleaning (described in the Analytic Issues section) was done, and the information is presented in the table below. This final cleaning shows that, after all other cleanings were finished, more students were being eliminated as a percentage at the lower grade levels than at the upper grade levels.

Grade Level	2	3	4	5	6	7	8
Sample Size Before Mobility Cleaning	5521	5878	5700	5343	5393	5138	5106
Final Sample Size	4774	5247	5279	4822	4978	4763	4798
Percent Decrease	13.5	10.7	7.4	9.8	7.7	7.3	6.0

After the cleanings, Rasch measures were assigned using the results of our equating study, and Rasch gains from 1988 to 1989 were calculated. In order to determine the predicted scores, least squares regressions were then run for reading and math separately at each grade level. For example, for the background index the students' gain was regressed against the following 'effects coded' variables: Black Effect, Hispanic Effect, Asian Effect, Gender Effect, and Old For Grade Effect. Two-way and three-way interaction models were run, but very few significant interactions were found, let alone systematic patterns of interactions across grade levels or within subjects, and therefore only main effect models were employed. The only exception to this was a number of interactions detected in seventh grade math, but this will have to be investigated further, and alone was not enough to warrant the use of interactions for that one level and subject. The results of these regressions will be discussed shortly.

Having obtained the regression coefficients, the cleaning procedures described above were then followed for all students in the system (again, roughly 25,000 to 27,000 per grade, with generally higher enrollments at the lower grade levels). In this way, actual gains were computed for all students who passed the cleanings, predicted gains were calculated using the regression results, and these gains were averaged to the school level. Finally, these school average gains were used to form the grade level and whole school indices for each school.

Results

Three findings based on the regression equations of the back-ground index were immediately apparent. First, as is shown in Figures 1 to 5 in Appendix A, the significance of the background index variable effects in general decreases across the grade levels for all of the variables. These effects become less 'explanatory' at the later grades.

Second, upon examining specific effects, it can be seen that certain patterns emerge. Figure 1 shows that Blacks in Chicago have lower gains, almost all of them being significant, in reading and math gains for all grade levels. This would tend to suggest that Blacks start lower in reading and math, and fall further behind their White counterparts (coded -1 in all of the race variables) every year as they continue in school. Frank and Seltzer, 1990, showed this phenomena in reading using Rasch scaled longitudinal data from the Chicago Public Schools. The Hispanic Effect variable (Figure 2) is statistically significant only at second and fourth grade for reading, where it can be seen that Hispanics make greater gains than Whites. Using longitudinal data from Chicago, Frank and Seltzer (1990) found that Hispanics started behind Whites in reading in first grade and then learned at a faster rate, nearly catching up to Whites by the sixth grade. Our cross-sectional data tends to support this hypothesis. However, the same pattern is not observed in math since there are no significant effects at any grade level. Except for second grade gains in reading, Asians gain more than their White counterparts in reading and math across all grade levels, most of these effects being statistically significant (Figure 3).

Figure 4 shows that being "old for grade" (coded as positive) tends to have a negative effect on gains, especially in the second grade where the effect is the largest, and especially in mathematics. This will be further investigated using the 'current retention' variable of the program index when that index is completed. Only about half the Gender Effect coefficients, coded positive for females, are significant at the .10 level, and those that are tend to show females making slightly greater gains (Figure 5).

Third, the R-squared values for the gain measure regressions (or the amount of variance in gain that the background index variables account for) for reading and math, for all grade levels, is extremely low, as shown in Figure 6. We believe that the reason for the low amount of variance accounted for is that, in using gain scores as our outcome, we are effectively controlling for the students' previous achievement, which has been shown in many studies to be the single best predictor of present achievement. Therefore, as a check of this hypothesis, we ran the regressions for the second index against the students' status (i.e., their 1989 Rasch measure), and indeed, the R-squared values are much higher, ending up in the .18 to .20 range in eighth grade where we might expect them to be based on much previous research using these types of variables. Thus, these background variables are more highly correlated with status than they are with gains in achievement, and we will have to consider this in our attempts to find correlates of school improvement. However, we still feel that taking these effects into account, even though they are small, can be seen as being better than not adjusting scores (as in the system index) for any background information at all.

Next, in order to calculate scores for the system index, achievement gains in reading and math had to be averaged across the whole system. These average system gains also showed some interesting patterns (Figure 7). First, the second grade gain for reading and math was quite a bit higher than all other grade levels. This is the same pattern that is observed with the Degrees of Reading Power exam, another Item Response Theory scaled standardized test, wherein the most growth in learning is shown to occur at the lowest grade levels.

Second, the amount of gain tends to decrease across the grade levels. Because of this we believe that, unlike the GE metric which forces growth to be the same across all grades by definition, the Rasch metric provides a more accurate picture of student learning across the grades. That is, we would expect the most growth in reading and math as students acquire the most basic skills in the lower grades.

Third, one can observe a "sawtooth" pattern in the average gains across the grade levels. For example, for reading and for math, the gains in second grade are higher than those in third, which are lower than those in fourth grade, which are higher than those in fifth, etc. An explanation for this can be found in the fact that these are gains, and therefore the post-test for one-year becomes the pre-test for the next. Thus, considering a given score for a first grader, a high score in second grade would imply a large gain for that grade, but then this high score would lead to a lower third grade gain. This 'sawtooth' pattern is also seen in examining individual schools' gain patterns, and will be discussed in the Results section using some examples. This autocorrelative feature of the use of gain measures is confirmed by previous analyses done by the Department of Research, Evaluation, and Planning (DREP) of the Chicago Public Schools.

Fourth, for reading and for math, much lower average gains are shown in the third grade, and relatively higher gains in the fourth grade. Previous work by the Chicago Public Schools on reading growth in Chicago using GEs had shown a "fourth grade dip," and DREP and the schools entertained many explanations for this phenomena. One theory held that the fourth grade ITBS test included a greater emphasis on higher order comprehension skills that students were unprepared for. Another theory suggested that fourth grade teachers taught very differently from lower primary grade teachers and that students had difficulty in making this transition. Our results actually show larger gains in fourth grade, but a dip in third grade. Smith (1992) shows this same pattern at third grade in Rasch scaled longitudinal mathematics data from Chicago. In addition, average gains for math are slightly higher in second grade than for reading, and quite a bit higher than in grades three through six. However, without better student background data and more exact information concerning the structure of the ITBS test and types of instruction used in schools at these grade levels, no hypotheses can be put forward as to why students seem to be gaining less in third grade, and why they are progressing more rapidly early on in math.

Next, after we had formed our school scores, we sought to find out how correlated they were within schools. This was an attempt to see if general patterns of lower or higher gains within a school could be seen in an effort to determine how improving schools might look. Therefore, correlations were run, separately for reading and math, to determine the

relatedness of gains across the grade levels. The results are given, for the background index, in Appendix B.

There are few significant relationships for both reading and math, in gains between the grade levels. Of the 21 correlations possible for grades two through eight, only six for reading and four for math are significant at the .10 level, and many of the significant gains are negative meaning that a higher gain in one grade level actually relates to a lower gain at another level. In general, this indicates that gains in achievement made at one grade in a school, as measured by the ITBS, have very little to do with gains made in another grade in that school. Moreover, it can be seen that correlations on the main diagonal, for reading and math, are for the most part negative, showing once again the 'sawtooth' pattern of gains where the gain in one grade level is negatively related to the gain in an adjacent grade. These two observations will now be further investigated using specific examples of school scores.

Sample School Results

We will now consider the actual results of four schools. The scores we will be referring to are contained in Appendix C. The first school, which we will call 'school A', is a small sized kindergarten through eighth grade school which is racially mixed having 47% White, 6% Black, and 47% Hispanic students. (All demographic information on these schools was obtained from the Illinois State Report Card data, and was not used in the statistical analyses presented here). Its whole school index scores, for the system and background indices for both reading and math, are very close to 100 indicating average achievement in this school as compared to all other Chicago Public Schools. (Note that almost all scores within a subject for the system and background indices are close to each other since the predicted scores do not differ much from the actual scores because the effects are all relatively small.) However, the grade level index score patterns are quite different. In reading and in math, better than average gains are achieved at second grade, very large gains in the third grade (counter to the systemwide pattern), and then very small gains are made in the fourth grade. It is likely that the large third grade gains led, at least partially, to the small fourth grade gains, and perhaps the instruction was poor (there is only one classroom per grade at this school), but without any further information about the exact scores of the students or the quality of instruction in that grade, an exact cause cannot be deter- mined. Also in school A, average fifth grade gains are followed by very low sixth grade gains, but again more information would be needed to attempt to establish a cause for these low gains. Therefore, one caveat we suggest in the use of our index scores is a need for further investigation as to why certain grades in a school might be faring poorly.

School B is a medium-sized magnet high school that also accepts high achieving (gifted) seventh and eighth graders. It is also racially mixed having 24% White, 56% Black, 8% Hispanic, and 12% Asian students. We would expect high scores for this school on all indices. However, this is not the case. Poorer than average gains are shown in both subject at both grade levels, and overall. The reason for this, after further investigation, was found to be that these students had "topped out" on the ITBS by the sixth grade, and therefore,

they did not have very much left that they could gain on the test in either seventh or eighth grade for reading or math. This could be what happened in the fourth grade in school A above. Therefore, a second caveat in the use of these indices is that these tests imperfectly measure achievement and achievement growth, and high achieving students might actually show less improvement than low achieving students because of the possibility of these ceiling effects.

Next we chose a medium-sized "racially identifiable" school (100% Black) that has very high mobility, high truancy, and a high percentage of students from low income families (92% of students participating in free lunch programs). The whole school indices for 'school C' are all quite low. The grade level indices are also all very low with the exception of the eighth grade, and they show that the students on average in reading in the third grade scored lower than they did in second grade (hence the negative index values). At least part of the reason for this is that many of these students in the second grade took a test that was one level below the level that most second graders take, formerly a common practice of teachers whose students were not prepared for the 'on level' test. Then these students took the correct test (level 9) in the third grade and scored very poorly. Perhaps many were confused by the harder exam as many seem to have failed to complete the test. In 1990, the Chicago Public Schools implemented a policy whereby all students in the system must take the 'proper' test for their grade level (we cannot be sure yet how closely this is being followed). However, a third caveat is that this policy was not implemented pre-reform, and the skipping of test levels (which happened almost exclusively at first, second, and third grades) seems to affect the grade level indices, even given that the tests have been adequately equated. However, the generally low gains in this school (and in other similar schools) are not accounted for by our system or background index models. This will be further addressed in the next section.

School D is a large racially identifiable school having 99% Hispanic and 1% Black students, also with very high poverty and mobility rates, in which 97% of the students are in a Bilingual program of some type. This is especially common at the lower grade levels, and no students in this school in second or third grade passed through our cleaning for being in a bilingual program three years or less (a Chicago Public School criterion). Even so, nearly all of the whole school and grade level index scores are at or above 100, and much less of a 'sawtooth' pattern is observed. However, again, other information on this school would be needed to attempt to explain this pattern of achievement, and thus our final caveat is that all information of this type, background information, program information, and indeed test scores themselves, might only help explain part of 'improvement' at the student or school level. This caveat is also further addressed in the next section.

Next Steps

Our next step will be to include the 1987 to 1988 gains to establish a more reliable two-year baseline for our growth indices. The program index scores will be also computed for these years, and then the 1989 to 1990, and 1990 to 1991 index scores can be computed because that data is already available.

In order to further determine the correlates of school improvement, once we have identified schools that make more or less gain than predicted according to our indices, we will try to find other school characteristics related to these gains. We intend to investigate other programs or services that schools offer (beyond those in the program index), school level budget allocations, staffing patterns, and other information available from central sources. In addition, we intend to use the information assembled from a recent survey of Chicago Public School teachers on their attitudes and instructional practices, a survey of principals about to be conducted on their attitudes, and priorities, and a possible small scale survey of our own to measure attitudes of students in selected schools as a complementary outcome to their achievement. A further targeted observational study of a small sample of schools is also possible.

Next, the basic data structures that we are working with are multi-level: achievement outcomes for students within classrooms within schools. Therefore, we intend to use Hierarchical Linear Modeling (HLM) an analytic tool in the future to help separate out and analyze effects at the different levels. Although class- room level data may not be a major factor for most analyses because of a lack of complete information, we expect to be using it for some auxiliary investigations. Moreover, a new HLM package is being developed that will utilize more than just two data points at a time (for students remaining in the same school over time), which is all that our current least squares gains regressions are able to do. This will be an extension of existing HLM models that will take into account the autocorrelative properties of gains discussed above. Thus, our basic multi-level structure will be the gains of students over grade levels within each school.

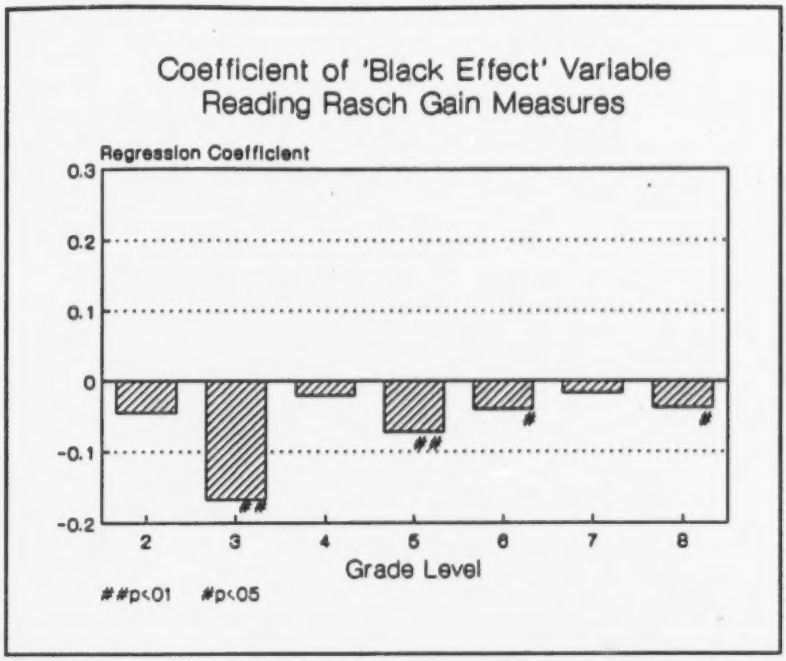
The problems involved in analyzing such multi-level data have been well documented (Burstein, 1980). Recent developments of HLM (Bryk and Raudenbush, 1991), also know as multi-level models (see, for example, Goldstein, 1987 and Bock, 1988) make more appropriate analyses of these data feasible. Therefore, we envision using these methods to: a) adjust for the measurement error and misfit of Rasch scaled achievement measures to disattenuate the 'masking' these errors have on other effects, b) develop better prediction equations for the denominators in our background and program indices based on the pooled within relationship, $ß_W$, rather than the ordinary least squares coefficients, $ß_T$, and c) model other possible sources of variability, that is, correlates of school improvement.

Finally, we hope to obtain two other sources of student level data that we believe will be correlated to student and school improvement. The first, as mentioned above, is student level mobility data. Using the Chicago Public Schools' "student history records", it is possible to determine the number of schools that a students attended in any year (up to six), and the percentage of time spent in each school. However, these records are unwieldy in their size and complexity, and will require much more work to be of use to us. (For a study of the effects of mobility on a longitudinal sample of students using these records see Kerbow, 1992.) Second, the 1990 census data should be available quite soon, and in this way we will be able to get data about family income, economic status, family educational resources and expectations, etc., at the block group level (about 1000 persons). Although this data is not available at the individual level, it should provide excellent information about the communities where students live.

One last note, the Illinois State Superintendent of Education has very recently revealed a plan that would put into place new methods of reporting achievement at the school level using that year's test scores (status), but also measures of improvement over the previous year, including such variables as change in individual test scores, change in school drop-out rates, and increasing or decreasing mobility rates. We applaud this effort, believing, as we have shown, that status alone, which is almost always the sole method of reporting, is often unfair as an accountability measure, and cannot fully explain learning at the student or school level.

APPENDIX A

Figure 1


Coefficient of 'Black Effect' Variable
Reading Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05

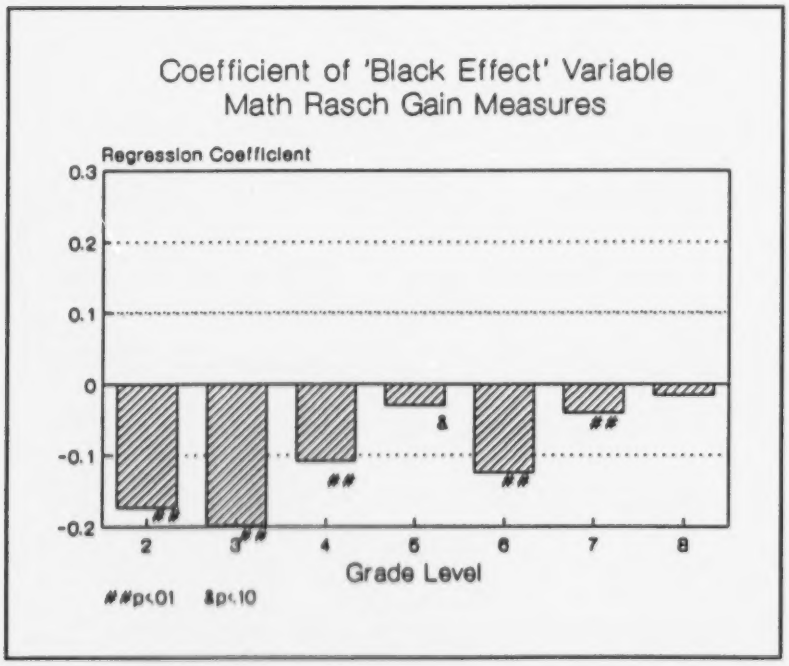
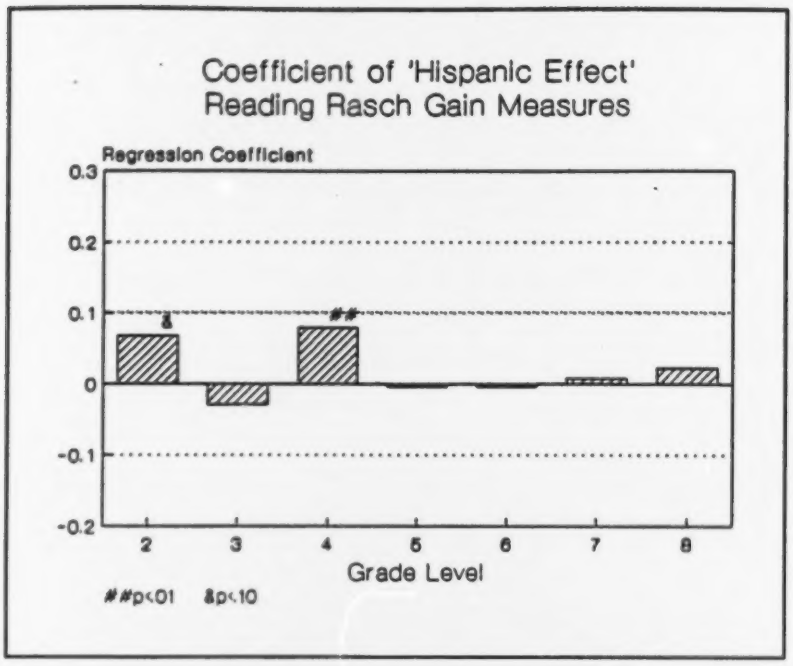

Coefficient of 'Black Effect' Variable
Math Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
2
3
4
5
6
7
8
Grade Level
##p<01 &p<10

Figure 2


Coefficient of 'Hispanic Effect'
Reading Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
&
##
2
3
4
5
6
7
8
Grade Level
##p<01 &p<10

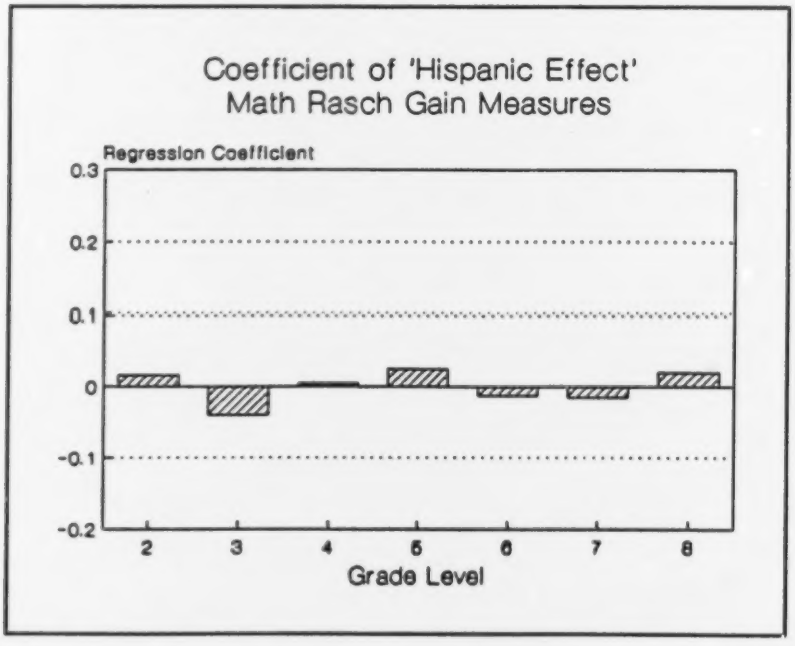
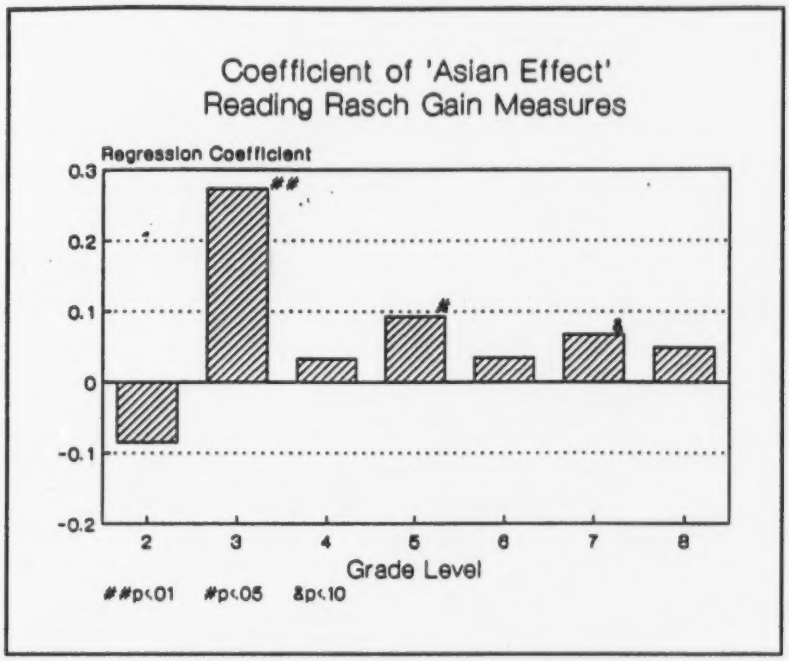

Coefficient of 'Hispanic Effect'
Math Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
2
3
4
5
6
7
8
Grade Level

Figure 3


Coefficient of 'Asian Effect'
Reading Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
##
#
&
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05 &p<10

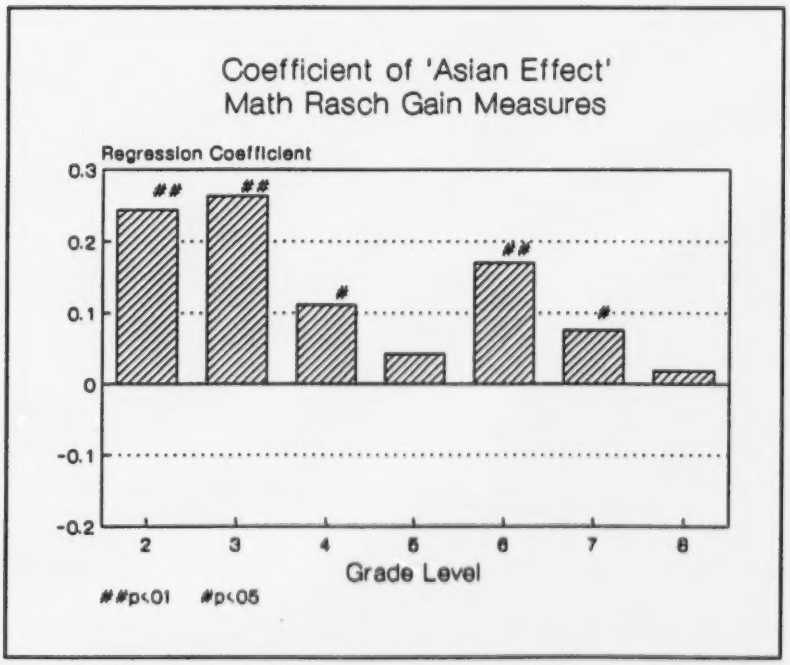
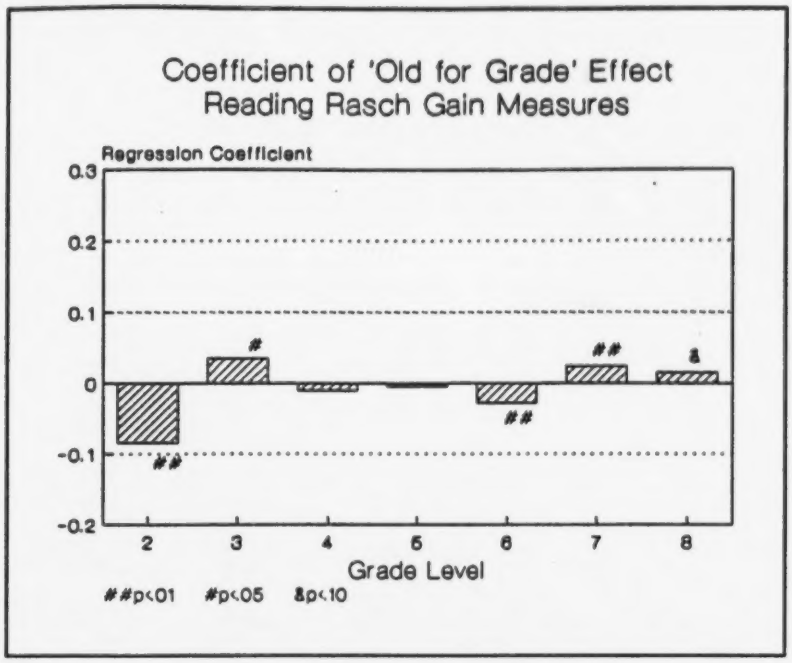

Coefficient of 'Asian Effect'
Math Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
##
##
#
##
#
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05

Figure 4


Coefficient of 'Old for Grade' Effect
Reading Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05 &p<10

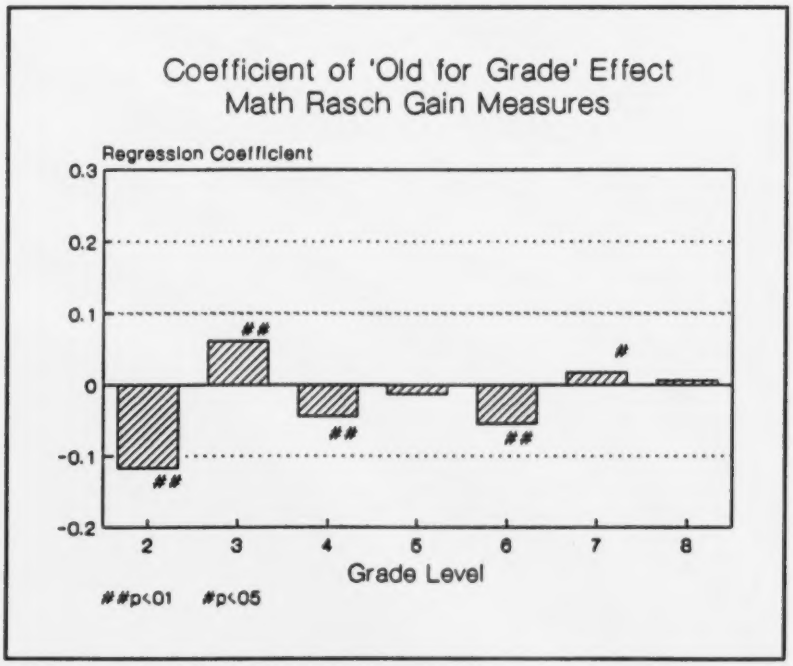
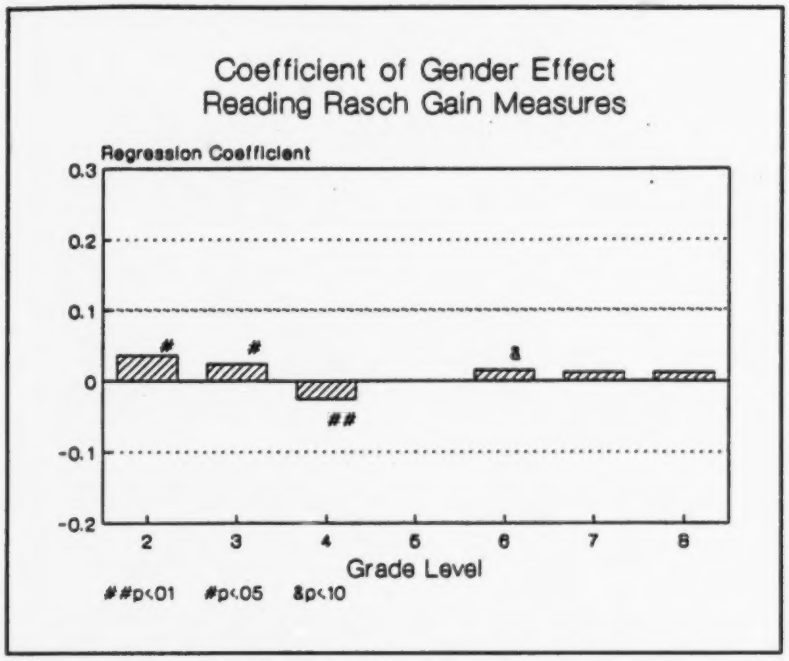

Coefficient of 'Old for Grade' Effect
Math Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05

Figure 5


Coefficient of Gender Effect
Reading Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
#
#
##
&
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05 &p<10

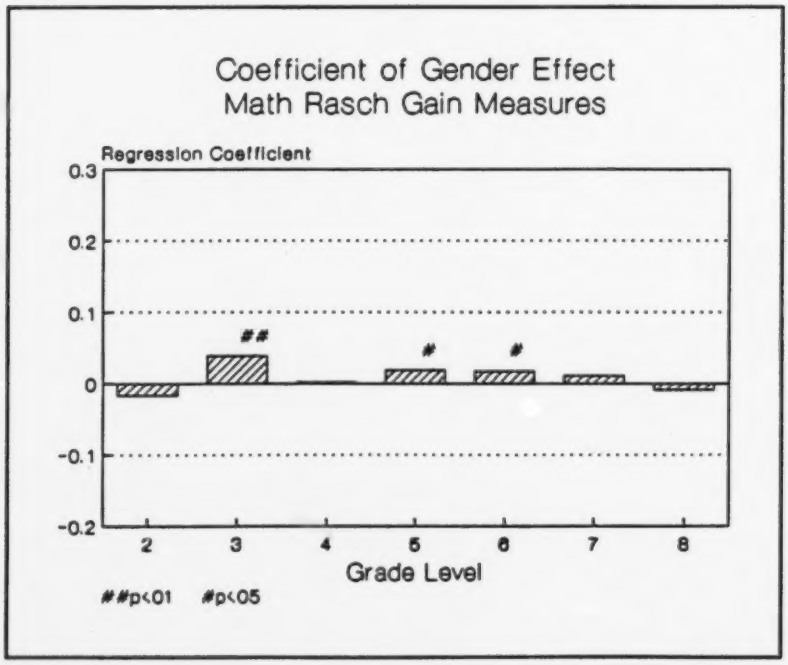
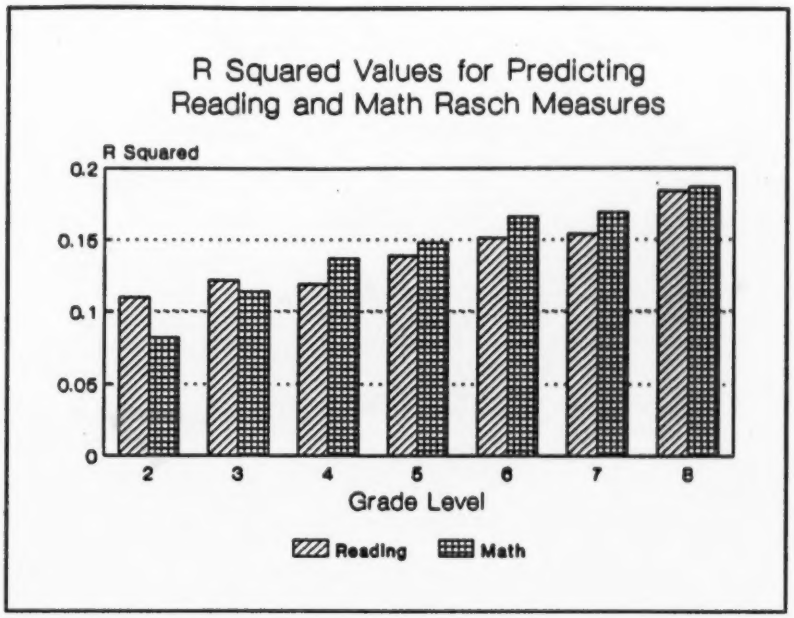

Coefficient of Gender Effect
Math Rasch Gain Measures
Regression Coefficient
0.3
0.2
0.1
0
-0.1
-0.2
##
#
#
2
3
4
5
6
7
8
Grade Level
##p<01 #p<05

Figure 6


R Squared Values for Predicting
Reading and Math Rasch Measures
R Squared
0.2
0.15
0.1
0.05
0
2
3
4
5
6
7
8
Grade Level
Reading
Math

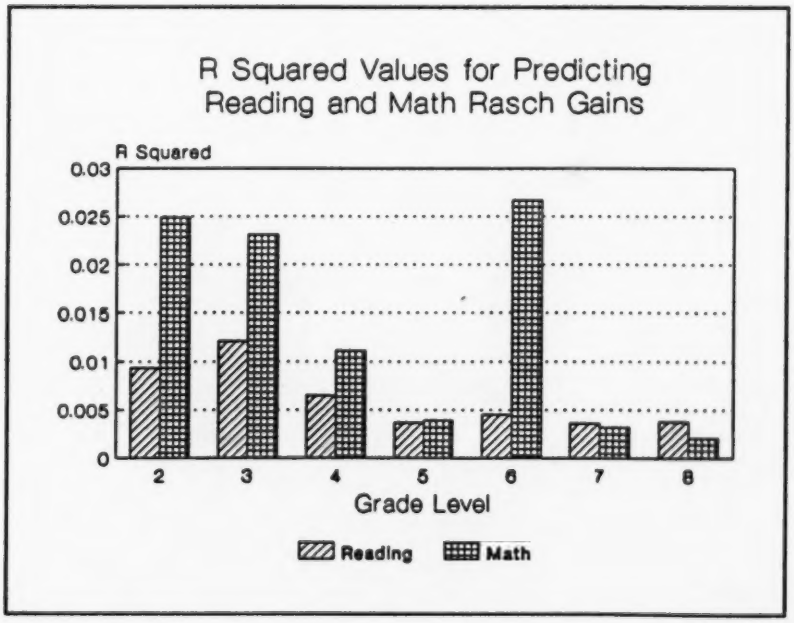
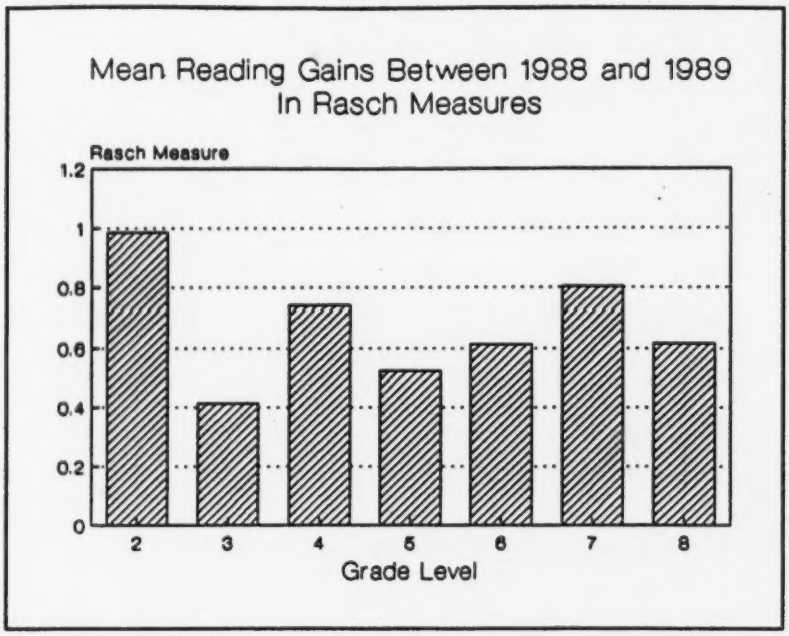
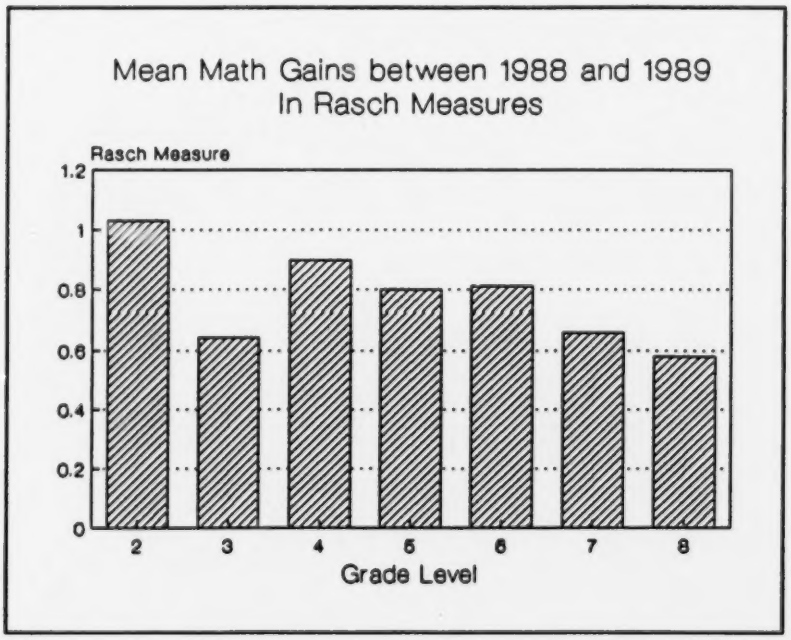

R Squared Values for Predicting
Reading and Math Rasch Gains
R Squared
0.03
0.025
0.02
0.015
0.01
0.005
0
2
3
4
5
6
7
8
Grade Level
Reading
Math

Figure 7


Mean Reading Gains Between 1988 and 1989
In Rasch Measures
Rasch Measure
1.2
1
0.8
0.6
0.4
0.2
0
2
3
4
5
6
7
8
Grade Level


Mean Math Gains between 1988 and 1989
In Rasch Measures
Rasch Measure
1.2
1
0.8
0.6
0.4
0.2
0
2
3
4
5
6
7
8
Grade Level

APPENDIX B

School Level Correlation Analysis -- Reading

	Second Grade	Third Grade	Fourth Grade	Fifth Grade	Sixth Grade	Seventh Grade
Third Grade	-0.035 (423)					
Fourth Grade	0.008 (416)	-0.093# (421)				
Fifth Grade	-0.109* (408)	-0.018 (413)	-0.035 (423)			
Sixth Grade	0.106* (390)	-0.123* (392)	-0.084# (402)	-0.005 (405)		
Seventh Grade	0.046 (360)	0.012 (363)	0.190** (372)	-0.022 (374)	-0.001 (384)	
Eighth Grade	0.000 (358)	0.043 (360)	-0.037 (369)	0.011 (372)	0.036 (380)	0.006 (388)

School Level Correlation Analysis -- Math

	Second Grade	Third Grade	Fourth Grade	Fifth Grade	Sixth Grade	Seventh Grade
Third Grade	-0.125* (423)					
Fourth Grade	0.055 (416)	-0.041 (421)				
Fifth Grade	-0.027 (408)	-0.058 (413)	0.010 (423)			
Sixth Grade	0.025 (390)	-0.065 (392)	0.081 (402)	0.149** (405)		
Seventh Grade	-0.034 (360)	-0.035 (363)	0.040 (372)	0.041 (374)	-0.050 (384)	
Eighth Grade	-0.005 (358)	-0.028 (360)	-0.015 (369)	0.110* (372)	0.018 (380)	0.109* (388)

** = p<.01 * = p<.05 # = p<.10

NOTE: The top number in each cell is the correlation coefficient, followed by the appropriate significance level of that number, and the number in parentheses is the number of schools used in each calculation.

APPENDIX C

The information below is refered to in the 'Sample School Results' section of the paper.

Number of Students Passing the Cleanings

School	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
A	31	35	32	30	27	45	28
B	0	0	0	0	0	75	66
C	49	61	46	56	44	43	49
D	0	0	129	99	107	116	124

System Index Grade Level Scores for Reading

School	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
A	141.1	235.2	45.8	113.5	62.2	93.7	127.8
B	.	.	.	.	.	48.4	93.6
C	73.7	-31.3	60.2	98.3	83.2	73.7	110.2
D	.	.	118.9	98.1	109.2	116.5	110.7

Background Index Grade Level Scores for Reading

School	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
A	135.7	213.2	47.7	106.8	59.8	97.9	126.3
B	.	.	.	.	.	49.8	94.9
C	77.5	-34.9	61.5	103.7	84.8	74.0	114.2
D	.	.	107.3	91.5	106.6	113.4	103.7

System Index Grade Level Scores for Math

School	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
A	106.3	187.9	89.9	93.3	83.5	48.4	89.0
B	.	.	.	.	.	70.7	57.8
C	48.3	52.9	79.6	56.7	63.0	75.6	118.3
D	.	.	137.6	104.5	128.2	90.5	166.2

Background Index Grade Level Scores for Math

School	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
A	103.0	166.0	84.6	92.8	79.0	48.3	88.4
B	.	.	.	.	.	71.4	58.5
C	52.6	58.8	84.5	58.0	66.1	77.0	120.0
D	.	.	127.7	99.8	119.3	88.6	158.9

Whole School Index Scores for Reading and Math

School	System Index Score for Reading	Background Index Score for Reading	System Index Score for Math	Background Index Score for Math
A	118.2	113.8	98.6	93.3
B	69.6	70.9	64.7	65.4
C	63.8	65.5	69.7	73.0
D	111.3	104.9	126.8	120.2

REFERENCES

Bock, D. E. (Ed.) (1989). Applications of multilevel models. New York: Academic Press.

Bryk, A.S. and Raudenbush, S. (1991). Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods.

Burstein, L. (1980). The Analysis of Multi-level Data in Educational Research and Evaluation. Review of Research in Education, 8, 158-233.

Easton, J. Q., Dean, P. and Bryk, A. S. (1991). School Change and Its Correlates: Studying Chicago Public Elementary Schools After School Reform. Paper presented to the annual meeting of the American Educational Research Association.

Frank, K. and Seltzer, M. (1990). Modeling Growth in Reading Achievement. Paper presented to the annual meeting of the American Educational Research Association.

Goldstein, H. (1987). Multilevel Models in Educational and Social Research. New York: Oxford University Press.

Kerbow, D. (1992). Networks of Exchange: The Consequences of Student Movement in Inner-City Schools. Paper presented to the annual meeting of the American Educational Research Association.

Lee, O. K. and Wright, B. D. (1992). Mathematics and Reading Test Equating. Paper presented to the annual meeting of the American Educational Research Association.

Schulz, E. M., Shen, L. and Wright, B. D. (1990). An Equal-interval Scale for Studying Reading Growth. Paper presented to the annual meeting of the American Educational Research Association.

Smith, J. (1992). Math Longitudinal Study. Paper presented to the annual meeting of the American Educational Research Association.